LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-673806

Large-scale First-Principles molecular dynamics simulationswith electrostatic embedding:

application to acetylcholinesterase catalysis

J. L. Fattebert, E. Y. Lau, B. J. Bennion, P. Huang, F. C. Lightstone

June 25, 2015

Journal of Chemical Theory and Computation

Disclaimer

Large-scale First-Principles molecular dynamics simulations with electrostatic embedding: application to acetylcholinesterase catalysis

Jean-Luc Fattebert,*,† Edmond Y. Lau,‡ Brian J. Bennion,‡ Patrick Huang,‡ and Felice C. Lightstone‡

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA, and Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA

E-mail: fattebert1@llnl.gov

Abstract

Enzymes are complicated solvated systems that typically require many atoms to simulate their function with any degree of accuracy. We have recently developed numerical techniques for large scale First-Principles molecular dynamics simulations and applied them to study the enzymatic reaction catalyzed by acetylcholinesterase. We carried out Density functional theory calculations for a quantum mechanical (QM) sub-system consisting of 612 atoms with an $O(N)$ complexity finite difference approach. The QM sub-system is embedded inside an external potential field representing the electrostatic effect due to the environment. We obtained finite temperature sampling

*To whom correspondence should be addressed
†Center for Applied Scientific Computing
‡Physical and Life Sciences

by First-Principles molecular dynamics for the acylation reaction of acetylcholine catalyzed by acetylcholinesterase. Our calculations shows two energies barriers along the reaction coordinate for the enzyme catalyzed acylation of acetylcholine. The second barrier (8.5 kcal/mole) is rate-limiting for the acylation reaction and in good agreement with experiment.

Introduction

Molecular dynamics (MD) is a widely used method in computational biology. Most often, classical potentials are used and interactions between atoms are simply described by elementary rules based on distances and angles between atoms. However, very important chemical phenomenon, such a bond breaking and forming, require accurate quantum mechanical (QM) models and cannot be fully described by classical potentials.

To avoid the high computational cost of representing the whole system of interest at the QM level, one can restrict QM modeling to a subset of atoms around the chemical reaction of interest and treat the remaining system with a more approximate model, either continuum or using classical particles. Combined quantum mechanics/molecular mechanics (QM/MM) methods are very popular for studying enzymatic reactions.[1,2] QM/MM comes in many flavors, depending on the QM approach, the MM force field, their coupling, and the termination of the bonds cut at the QM/MM boundary. The size of the QM region obviously matters as charge transfer and polarization effects in the vicinity of the region of interest can have a large influence on the phenomenon of interest. Thus, fairly large QM regions (150–1000 atoms) are required to achieve reliable accuracy.[3,4] Liao and Thiel found that a QM region of 408 atoms was required to get converged energies for the reaction catalyzed by acetylene hydratase.[5] Hu et al.[6] recently carried out calculations with QM regions including as many as 930 atoms to study the reaction energies for hydrogenase.

Even with a restricted subset of atoms at the quantum level, QM/MM molecular dynamics is limited to a few picoseconds, which is insufficient for accurate sampling of the available

phase space for complex biological and chemical molecular systems. Often classical MD is performed to obtain sampling over a longer time-scale, and QM calculations are carried out on a limited number of snapshots.[6,7] Considering large QM regions however introduces new problems, if geometry optimization is to be used for these structures a large number of local conformational minima are introduced.

In this paper, we are particularly interested in evaluating the free energy barrier between two given states in a chemical reaction. If one has a good guess of the chemical reaction coordinate, thermodynamic integration can be used to compute the free energy along that path.[8] Often a simple reaction coordinate can be chosen as a scalar parameter λ to parameterize that path. We calculate the free energy difference by integrating over energy changes along the path,

$$F(\lambda_2) - F(\lambda_1) = \int_{\lambda_1}^{\lambda_2} \langle \frac{\partial U(\lambda')}{\partial \lambda'} \rangle_{\lambda'} d\lambda' \tag{1}$$

where U denotes a parameterized potential energy and $\langle \dots \rangle_{\lambda'}$ denotes a statistical average over the equilibrium ensemble associated with the parameter value λ'

In practice, we perform this calculation by sampling a series of equilibrium configurations for a finite set of values of λ', using the multiconfiguration thermodynamic integration approach proposed by Straatsma and McCammon.[9] For each value of λ', First-Principles Born-Oppenheimer molecular dynamics in the NVT ensemble is used to obtain an ensemble average of the derivative of the energy of the system with respect to λ. Finally, we obtain the free energy by numerically integrating the derivative in Eq. (1). Other authors have used similar thermodynamic integration techniques to compute free energy profiles in biomolecular reactions using *ab initio* MD.[10]

In our numerical approach, we obtain our ensemble average by carrying out First-Principles molecular dynamics (FPMD) for a subset of atoms comprising the biological system of interest (QM region), embedded in a frozen electrostatic environment. We use an electrostatic embedding model similar to the one commonly used for QM/MM, but instead of including charges from MM, we pre-calculate a static external potential which we

use throughout our molecular dynamics simulation. The electrostatic potential outside of the active site likely does not change substantially during equilibrium dynamics. We calculate this potential using charges from the initial force-field model, corresponding to atoms not included in the QM region, as proposed by Fattebert et al.[11] We use our recent electrostatic embedding model in combination with an $O(N)$ complexity First-Principles molecular dynamics approach[12,13] to calculate energy barriers in a chemical reaction with a large QM region, involving hundreds of atoms. We describe the reaction path by reaction coordinates corresponding to distances between atoms and distance differences between two pairs of atoms. We use constraints on reaction coordinates to sample the reaction path, and we integrate the average forces required to enforce these constraints to obtain energy barriers.[14] Fox et al.[7] have also used $O(N)$ DFT methods for large-scale first-principles calculations. Contrary to our approach though, these authors use a classical potential for MD sampling and performed QM calculations only on a finite number of snapshot configurations.

Embedding in an external electrostatic potential is particularly important when a QM region contains a net charge. While Linear Combination of Atomic Orbitals (LCAO) methods restrict the electronic wave functions within the vicinity of the QM atoms by an artifact of the basis set, more general approaches such as the Finite Differences method we used in this study or the Plane Waves approach do not confine electrons. While this absence of artificial confinement is a desirable feature from a general numerical point of view, this can lead to various problems if the net charge of the QM system is incorrect.

Using the methods described above, we chose to study the enzymatic reaction catalyzed by acetylcholinesterase. Acetylcholinesterase (AChE) plays a vital role in the central nervous system by hydrolyzing the neurotransmitter acetylcholine (ACh), resulting in termination of impulse signaling.[15] The catalytic rate of hydrolysis of ACh by AChE is one of the highest in nature and close to diffusion-limited (turnover number 7.4×10^5/min).[16] The active site is deeply buried within the enzyme, and using a combination of hydrogen bonds and hydrophobic regions positions ACh in close proximity to the catalytic triad of Serine203-

Histidine447-Glutamate334 (human numbering) for facile reactivity.[17] Acetylcholinesterse is among the most widely investigated enzymes by both experimental and computational methods. Of important interest is the binding and inhibition of AChE by organophosphates.[18,19]

The reaction of Ach catalyzed by this enzyme occurs in two stages, acylation and deacylation.[20] Transition state and intermediate structures are very difficult to study experimentally but are ammendable to computational studies. In this contribution we have used large scale QM/MD simulations to study the structures and energetics of the acylation reaction of ACh catalyzed by AChE (see Fig. 1) and compare our results with previous studies.

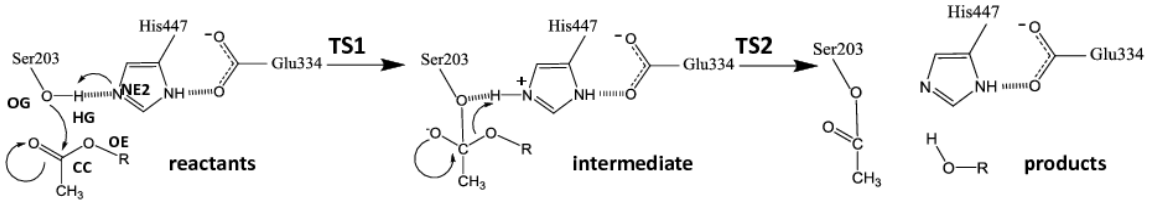

Figure 1: Acylation reaction of acetylcholine catalyzed by acetylcholinesterase.

Modeling and numerical methods

Molecular system preparation

We performed all classical molecular dynamics calculations using the program NAMD[21] with the CHARMM force field.[22] Acetylcholine was modeled into the active site of the apo human acetylcholinesterase structure (PDBID: 1B41)[17] using the program Autodock (version 3.05).[23] We solvated the protein in a TIP3P[24] water box (81.8 $\times$ 70.1 $\times$ 87.8 Å^3) sufficient in size to have at least 12 Å of water between the protein and interface. We added counter ions (Na^+ and Cl^-) to the solution to physiological concentrations to neutralize the system. The final system contained a total of 46879 atoms. We optimized the geometry of the system by energy minimization using 100 steps of the conjugate gradient algorithm. We performed molecular dynamics starting from the energy minimized system. The electrostatics was treated by particle mesh Ewald summation with a grid spacing of 1

Å.[25] Non-bonded interactions were cutoff at 12 Å, and a 2 fs time step was used for the simulation. The system was heated to 300 K and equilibrated for 1.2 ns. We used the final structure from this dynamics as the starting point for the QM calculations.

To perform our First-Principles calculations, we divided the atomistic system into two parts: a primary (QM) system, encompassing the active site and several of its neighboring residues and a secondary subsystem (the remaining atoms) which are not included in the QM calculation. All residues within 9 Å of the hydroxyl oxygen in Ser203 were considered the active site of the enzyme and included in the QM region (612 atoms, Figure 2). For the secondary subsystem, we generated a list of atomic positions and associated partial charges (using CHARMM parameters). Based on that list of point charges, we generated a charge distribution given by a linear combination of Gaussian charge distributions associated with each atomic point charge. The total charge of this distribution is equal to the sum of the point charges and is the opposite of the total charge of the QM system for an overall neutral system. We evaluated the sum of all these Gaussian charges at each node of a uniform mesh covering the physical domain. We used a mesh spacing of ≈ 0.12 Å, which results in a global uniform mesh of size $672 \times 576 \times 704$. We then calculated the Coulomb potential associated to that charge field by solving a Poisson equation on this mesh by finite differences, using a multigrid solver. We used the resulting Coulomb potential as an external potential for the QM calculations.

To prepare the QM system, we need to properly terminate the bonds cut in the process of partitioning the atomistic simulation into a primary and secondary subsystem. We followed the protocol described in our previous paper.[11] Briefly, partitioning of a QM residue occurs at both peptide bonds. The bond between the NH and CA within the residue was cleaved and a hydrogen was used to cap the CA to form a methylene group. We also cleaved the bond between the NH and CA in the trailing residue and capped it with a hydrogen at the NH to form an amide group. (see Fig. 3) Proper termination of the protein chain minimized disruption of the local electronic structure. Water molecules are either fully included in

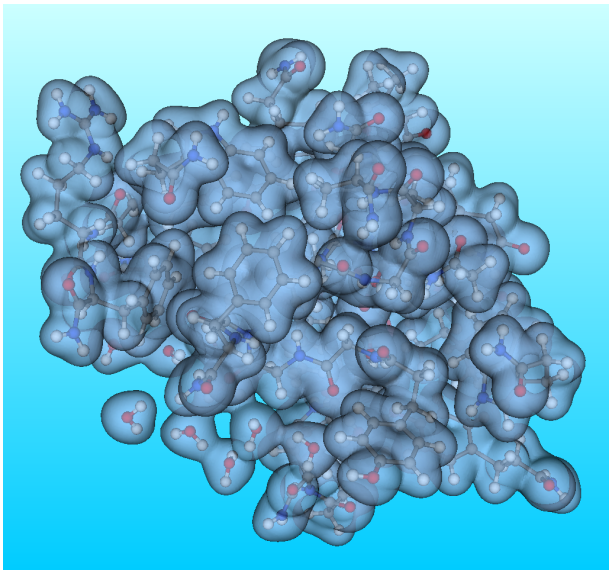

Figure 2: The QM region of the AChE active site QM region in the reactant state. The isosurface of the calculated electronic density is overlayed on the 612 atomic centers.

the QM region or not at all. The resulting QM region is composed of 612 atoms. Within our pseudopotential approximation and neglecting spin, we end up with 848 electronic wave functions to calculate.

Figure 3: Partitioning of a residue in our QM/MM scheme.

We carried out our quantum calculation on a fraction of the original domain used for the MM system. We used a QM domain of dimension $36.57 \times 30.82 \times 31.69$ Å^3, which corresponds to a subset of the mesh points used for the Coulomb solver. The resulting mesh covering the QM computation domain is of size $304 \times 264 \times 264$. Since the QM mesh is a subset of the global mesh, the external electrostatic field is defined on the QM calculation by a simple injection of the global Coulomb potential.

Once this external potential is set, we optimized the geometry of the termini atoms in

this Coulomb field with all the other atoms fixed. Finally, we lock the positions of a shell of atoms at the boundary of the QM region and let only the atoms in the inner regions to move freely, including the active site residues. This results in a total of 212 free QM atoms. Figure 2 shows an example of electronic density isosurface calculated for the resulting QM problem.

First-Principles molecular dynamics

Our First-Principles molecular dynamics approach is based on the $O(N)$ DFT methodology proposed by Fattebert and Gygi.[12,13] In that approach, for N doubly occupied states, the DFT equations are formulated in a general form without orthonormality constraints

$$\begin{aligned} E_{KS}[\{\phi_i\}_{i=1}^N] \quad &= \sum_{i,j=1}^{N} \left(S^{-1}\right)_{ij} \int_\Omega \phi_i(\mathbf{r}) \left(-\nabla^2\right) \phi_j(\mathbf{r}) d\mathbf{r} \\ &+ \frac{1}{2} \int_\Omega \int_\Omega \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 + E_{XC}[\rho] \\ &+ \sum_{i,j=1}^{N} 2 \left(S^{-1}\right)_{ij} \int_\Omega \phi_i(\mathbf{r}) (V_{ext}\phi_j)(\mathbf{r}) d\mathbf{r}. \end{aligned} \tag{2}$$

where the terms on the right side are the kinetic energy, the electrostatic energy, the exchange and correlation energy, and the potential energy of the electrons in the potential V_{ext} modeling the atomic cores, respectively. The electronic density is given by

$$\rho(\mathbf{r}) = 2 \sum_{i,j=1}^{N} (S^{-1})_{ij} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}). \tag{3}$$

We used the pseudopotential approximation in its Kleinman-Bylander form[26] and the PBE exchange correlation functional.[27] We discretized the energy functional (Equation 2) by finite differences on a uniform mesh. We confined each electronic orbital $\{\phi_i\}_{i=1}^N$ to a specific spherical region of radius R_c. These confinement regions limit the number of degrees of freedom associated with each orbital to a constant that does not increase with system size.

This leads to $O(N)$ degrees of freedom to represent N electronic wavefunctions, and an $O(N)$ computational complexity for the most computationally demanding operations. This truncation leads to an approximate solution with an error on the calculated atomic forces which decays exponentially fast as a function of the localization region radius.[12] We chose these regions to be large enough to reduce this error to an acceptable tolerance. For this study, we used $R_c = 4.23$ Å. This leads to average cutoff errors of 0.6 kcal/mol/Å for the atomic forces according to our benchmarks tests. Confined electronic orbitals are typically localized on a chemical bond, or where a lone pair would reside. Fig. 4 illustrates our approach by displaying isosurfaces of a few of these localized orbitals as obtained in our QM system.

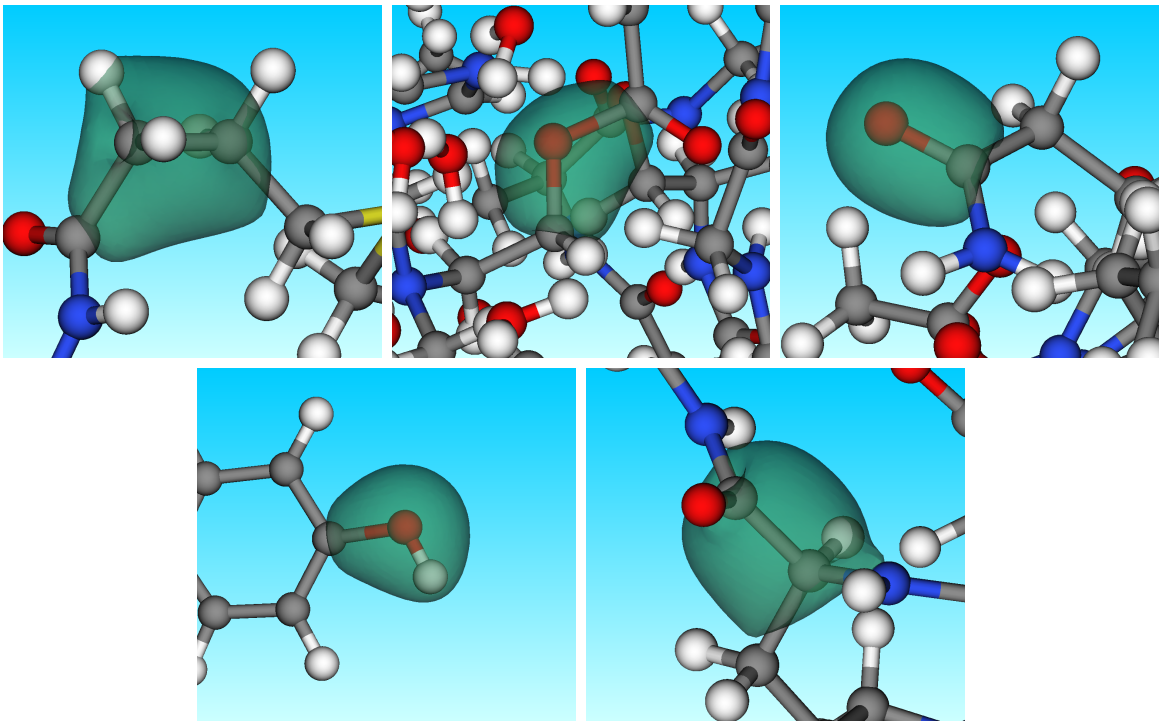

Figure 4: Isosurfaces of a few localized orbitals.

We used a Langevin thermostat to obtain a target temperature of 300 K. Deuterium masses were used for the hydrogen atoms, and normal masses for all other elements. A time step of 0.36 fs was used. In our approach, the electronic orbitals confinement regions evolve with the dynamics of the atoms. At every MD step, they are adapted based on the center

of charge of the latest computed orbitals.[13] We carefully monitored the amplitude of the displacement of these centers. When a charge center move is much larger than average, it typically means that the solution of the DFT equations fell from a local minima associated with the orbital truncations[13] — into another minima of lower energy. This is not a cause of concern in itself since those minima lead to very similar atomic forces. However, such changes reduce convergence of the electronic structure iterative solver. To remedy this slowdown issue, we added an extra adaptation step and an extra iterative cycle to solve for the electronic wave functions every time a move is larger than a set tolerance (0.026 Å).

Free energy barrier computation

As described in the introduction, we used thermodynamic integration to compute the free energy along the reaction path.[8,14] Acylation of Ach by AChE goes through two energy barriers with a tetrahedral intermediate in between.[28,29] In the first part of the reaction, the Ser203 (human numbering) hydroxyl oxygen (OG) approaches the carbonyl carbon (CC) of acetylcholine and bonds with it. Simultaneously, the hydroxyl hydrogen (HG) is transferred to the neighboring nitrogen atom in the imidazole (NE2) of His447. To sample the first part of the acylation reaction (reactants to intermediate), we use the OG—CC distance as the reaction coordinate and ran several simulations at fixed (constraint) values along the reaction coordinate.

In the second part of the acylation reaction, the formerly hydroxyl hydrogen (HG) on the NE2 of His447 is transferred to the ester oxygen atom of the acetylcholine (OE) resulting in bond breaking with the carbonyl carbon (CC). The reaction path is obtained by constraining the difference in distance between HG and the atom it is bonded with in the intermediate (NE2), and the distance between HG and the resulting hydroxyl oxygen that it is bonded with in the final choline product (OE). That is, the reaction coordinate is d(HG-NE2)-d(HG-OE). We sample this path by choosing a discrete number of values for that reaction coordinate and running independent simulations for each of these values as a constraint.

We imposed constraints between atoms using the SHAKE algorithm.[30] For each value λ of these constraints, we ran a molecular dynamics simulation at 300 K for at least 4 ps. We obtained out initial conditions by geometry optimizations under given constraints, producing the starting geometries of the reactants, products, and intermediate states. To calculate a free energy barrier with thermodynamic integration, the forces required to satisfy the constraints corresponding to λ must be determined. We calculated the average values of these forces over the last 3 ps of a simulation, considering the earlier time steps as equilibration time. Note that for the second reaction coordinate, we included a correction to the forces according to Sprik and coworkers.[14,31] We observe, however, that these corrections are negligible. Three picoseconds of sampling is typically not sufficient to converge statistical errors on an average of correlated data, but it does give some reasonable sampling of the configuration space for the QM system. It also provides enough time for the system to *forget* about possible biased initial conditions. We used the running average forces to estimate if the system has reached a stable regime on the time scale considered. If we observed a significant drift in the running average, we extended the simulation for a few extra picoseconds until the system settled into a more stable regime. We stopped the simulations when the 3 ps running average fluctuation was below a tolerance of 1.2 kcal/mol/Å over one picosecond. Figure 5 shows an example of such a behavior, that required over 6 ps to reach a non-drifting running average force.

Typically, the QM/MD simulations were carried out using 363 CPUs of a Linux cluster (Intel Xeon EP X5660). With this resource, a single MD step took about 100 seconds (wall-clock time).

Results and Discussion

The geometry for the reactant ACh, in the active site of AChE, is well positioned for nucleophilic attack by the hydroxyl oxygen of Ser203. Two hydrogen bonds are formed between

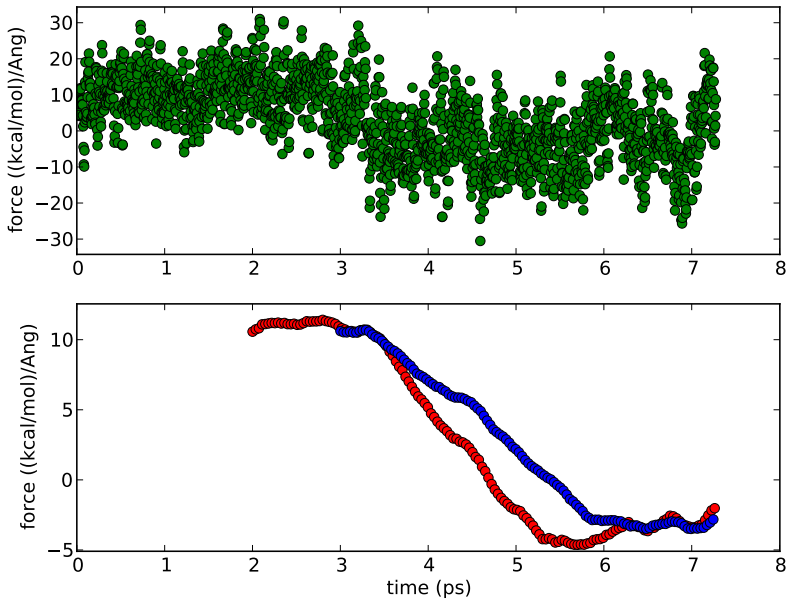

Figure 5: Illustration of instantaneous force acting to maintain constraint (Top) and the 2 ps (Red) and 3 ps (Blue) running average of that force (Bottom) as a function of time for the MD run closest in energy to TS2. Only every other ten points is shown for the instantaneous force and every other hundred for the running averages. The data show a shift in the force due to a slight change in the geometry towards a more product-like structure for the second transistion state.

the backbone amides (NH) of Gly121 and Gly122 (oxyanion hole) with the carbonyl oxygen (O) of ACh at distance of 2.93 and 2.79 Å, respectively. The average distance of OG of Ser203 to CC of ACh is 2.71 Å. The OG to NE2 of His447 distance is 2.69 Å which is reduced relative to the crystal structure distance of 2.90 Å. At the first transition state (TS1), the hydroxyl oxygen (OG) of Ser203 has lost its hydrogen (HG) to NE2 of His447. The HG-OG distance has increased from 1.01 Å in the reactant geometry to 1.47 Å in the TS1 structure. The loss of the hydrogen greatly enhances the nucleophilicity of the hydroxyl oxygen, and the separation between OG to CC of ACh is significantly reduced to 1.85 Å.

Based on the average forces measured on the constraints during the MD simulations, as well as the average values of the reaction coordinates in the reactants, intermediate and products states, an energy profile can be computed by a simple second order polynomial interpolation between data points along the reaction coordinates, see Fig. 6.

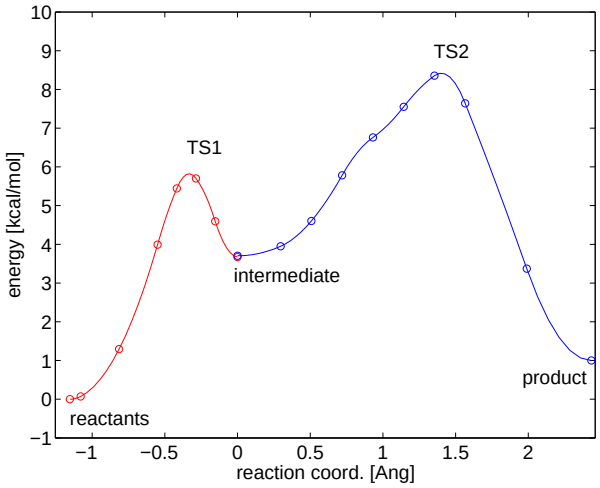

Figure 6: Energy barrier along the reaction coordinate. The rate determining step is predicted to be the second transition state with a barrier of 8.5 kcal/mol comparable to a value of 11.8 kcal/mol derived from experiment. The calculations also predict a slightly endothermic acylation reaction.

The free energy barrier for the first transition state is about 6.0 kcal/mol. This is lower than the activation barriers of 10.5 calculated by Zhang et al.[32] and 12.4 by Zhou et al.[33] A charged tetrahedral intermediate structure is formed after passage of the first transition state. This intermediate is formed between ACh and Ser203, and is 2 kcal/mole lower in

energy than TS1. Previous calculations have also found an intermediate structure that is 1-2 kcal/mole lower in energy than TS1.[32,33] The OG to CC distance in the intermediate is 1.56 Å. This charged intermediate is stabilized by the hydrogen bonds from the oxyanion hole that have almost identical distances to the reactants geometry. Additionally, a third hydrogen bond is formed between the formerly carbonyl oxygen (O) of Ach and the backbone amide of Ala204 to provide further stabilization of the extra negative charge. The amide NH of Ala204 to O of ACh distance contracts from 3.32 Å in the reactant structure to 2.82 Å in the intermediate structure.

If unconstrained, the intermediate is unstable at $T = 300$ K and its structure reverts to the reactants. It is necessary to fix the OG—CC distance at a value of 1.57 Å to sample this geometry in the MD simulation.

A second transition state (TS2) is formed when the imidazole of His447 rotates away from Ser203 and forms a hydrogen bond with the ester linkage of ACh (HG-OE distance 1.41 Å). This change in position is commonly seen in crystal structures in which Ser203 has formed a covalent adduct.[34,35] This new interaction causes the CC-OE bond in ACh to elongate from 1.50 Å in the intermediate structure to 1.91 Å in the TS2 structure. The TS2 structure is 4.5 kcal/mol higher in energy than the intermediate. The acylation reaction is complete when HG is transferred from His447 to the ester oxygen of ACh, causing the cleavage of the bond between CC and OE. These important distances are summarized in Table 1, and the active site geometries are shown in Figure 7.

We calculated the free energy barrier for the enzymatically catalyzed acylation of acetylcholine to be 8.5 kcal/mol (see Figure 6), which is in reasonable agreement with the experimental values of 11.8 kcal/mol estimated from the application of simple transition state theory[36] to the acylation ($1.7\text{x}10^6$/min) rate published by Froede and Wilson.[28] The limiting step for the acylation reaction from our calculations is the transition from the intermediate to product. This is in qualitative agreement with QM/MM calculations by Zhou et al.,[33] although quantitatively their activation barrier is larger at 13.2 kcal/mol. Additionally, Zhang

Table 1: Interatomic distances describing important interactions for the reactants (ES), first transition state (TS1), intermediate (INT), second transition state (TS2), and product (EP).

Distance Å	ES	rms	TS1	rms	INT	rms	TS2	rms	EP	rms
OG–CC	2.79	0.19	1.85	0.0007	1.56	0.0002	1.40	0.04	1.35	0.02
OG–HG	1.02	0.03	1.44	0.16	1.59	0.09	2.44	0.14	2.97	0.24
HG–NE2	1.70	0.12	1.15	0.09	1.07	0.03	1.17	0.02	1.90	0.16
HG–OE	3.11	0.20	2.67	0.11	2.58	0.16	1.33	0.02	1.00	0.02
CC–OE	1.35	0.03	1.43	0.03	1.50	0.04	2.13	0.30	3.04	0.19
OE(E334)–NE2	2.66	0.07	2.62	0.06	2.60	0.06	2.82	0.09	2.88	0.11
CO–N(Gly121)	2.95	0.10	2.95	0.08	2.92	0.11	3.15	0.17	3.27	0.14
CO–N(Gly122)	2.78	0.10	2.77	0.07	2.78	0.08	2.84	0.10	2.88	0.09
CO–N(Ala204)	3.35	0.15	3.02	0.11	2.85	0.09	2.86	0.11	2.88	0.11
O(E202)–HE2(H447)	4.57	0.23	4.45	0.26	4.90	0.23	4.60	0.21	4.18	0.23
O(E202)–H(Wat)	1.76	0.14	1.82	0.16	1.78	0.11	1.78	0.12	1.80	0.15
angle										
NE2-HG-OE	87		113		86		167		169	

and coworkers have shown that the activation barrier for the acylation reaction can vary by as much as 5 to 6 kcal/mol for different starting structures depending on the size of the QM system and the level of theory.[37]

The observed difference in free energies can come from many sources. It has been shown experimentally that distant charges from the active site can influence the binding affinity of the enzyme.[38] We expect our electrostatic embedding scheme, even if static, to handle charge effects correctly. However, DFT calculations tend to under estimate transition state barrier heights.[39,40] Additionally, the PBE functional underestimates reaction barriers relative to B3LYP.[41] In the calculation of the deacylation reaction, Zhou et al used two different functionals (B3LYP and PBE0). The PBE0 functional[42] is a hybrid version of PBE. The details of the reaction did not change, but the relative energies differed. The B3LYP transition state and product energies were 2.6 and 4.8 kcal/mole higher in energy respectively, than the corresponding PBE0 energies. Although molecular dynamics was used to sample the phase space, it may not have completely removed the bias associated with the starting geometry. However, this is still better than 0 K simulations where the optimized geometries strongly depend on the initial conditions, and forces required to maintain the constraints on

a given reaction coordinate can differ significantly between the various local minima.

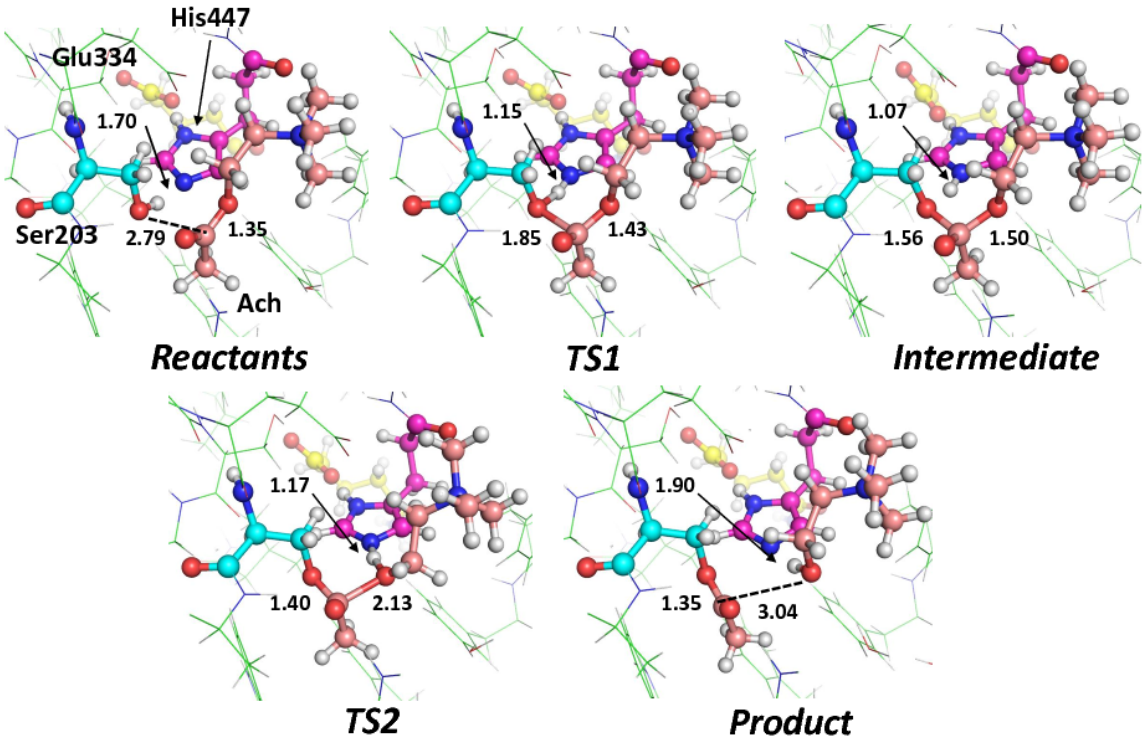

Figure 7: Active site geometries for the reactants, TS1, intermediate, TS2, and product. Atoms of the catalytic triad and ACh are shown in ball and stick representation and the carbon atoms are colored as follows: Ser203(Cyan), His447(Magenta), Glu334(Yellow), and ACh(Salmon). Select active site distances undergoing change are shown and the values are in angstroms. An animation of this reaction coordinate is included in the Supplementary Information.

A comparison of our acylation reaction coordinate and active site geometries to Zhou et al.[33] reveals similarities and several important differences. The geometries of the reactant, TS1, and intermediate states are in general agreement. The tetrahedral intermediate state is more stabilized in our calculations when compared to Zhou et al. This increased stability is due to the more product-like TS2 geometry (see Table 1). Unconstrained MD simulations of the tetrahedral intermediate readily transitions to the reactants which also suggests the TS2 is more product-like. This shift in the TS2 geometry to the product also effects the final energy of the acylation in that our calculations reveal a very slightly endothermic acylation reaction. However, the reaction coordinate of Zhou et al. reveals a significant exothermic

acylation reaction (-6.0 kcal/mol). We were unable to find experimental results to either support or refute the final energetics. However, once the choline is displaced from the active site only the forward reaction is possible. Our TS2 geometry as described by the distances shown in Table 1 are markedly different than those of Zhou. It is not clear what is causing the difference between the two reports. Although, inspection of the chosen reactions coordinates suggests that different restraints may be one possible source.

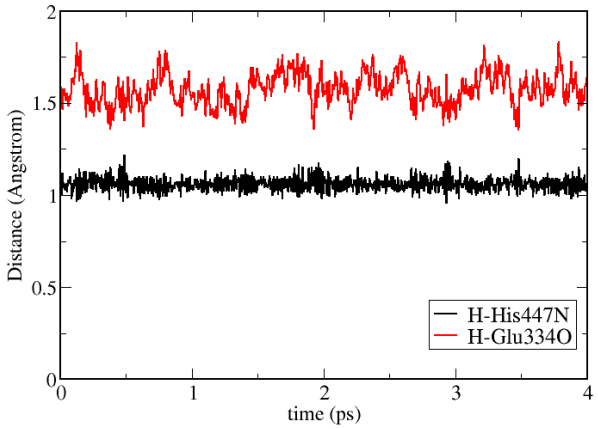

Figure 8: Hydrogen interaction between Glu334 carboxylate O and His447 NE1 atoms for the tetrahedral acyl intermediate. We show the H-N and H-O distances. Based on these data, there is no evidence of the hydrogen bond shifting between the N and O atoms of His447 and Glu334, respectively

The nature of the hydrogen interaction between Glu334 carboxylate O and His447 imidazole NE atoms has been debated in the literature.[43] We observe the hydrogen atom to be tightly bonded to the His447 NE1 nitrogen for the entire MD simulation (4 ps) of the tetrahedral acyl intermediate. This observation is in contrast to the results published by Zhou et al.[33] which show frequent sliding of the hydrogen between the Glu334 carboxylate O and His447 imidazole NE atom during their 30 ps MD simulation. The difference in results could be due to the functionals used for these calculations. The PBE functional is more accurate for calculating hydrogen bond strengths[44] and the B3LYP functional is known to underestimate hydrogen bond energies.[41]

Concluding remarks

In this paper, we presented a new numerical approach for calculating free energy barriers with large QM regions. We carried out large QM simulations (612 atoms) embedded in a static electrostatic field to model the acetylcholine acylation reaction catalyzed by AChE. Application of our first-principles calculations show general agreement with previous experimental and computational studies. The reaction barrier heights and rate limiting step of the acylation reaction are correctly predicted within the uncertainty of experiment and theory.

We compared our numerical results with recent calculations by Zhou et al.[33] We found some differences in our results, in particular the height of the energy barrier and the differences in geometry in TS2 which may be explained by the size of our QM region which is an order of magnitude larger than previous simulations.[33,37] In our model, the larger number of atoms allows for greater QM interactions from atoms surrounding the catalytic active site.

Although the accuracy of our approach is ultimately limited by the accuracy of DFT in this context, we conclude that our approach is quite robust. This technique will be soon applicable to larger system sizes with faster computers and advanced scalable $O(N)$ electronic structure methods.

Acknowledgement

We thank Eric R. Schwegler for helpful discussions. We thank the Defense Threat Reduction Agency for funding (CBS.SCIC.01.10.LLNL.004). We also thank Livermore Computing for the computer time. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

(1) Senn, H. M.; Thiel, W. *Angew. Chem.-Int. Edit.* **2009**, *48*, 1198–1229.

(2) Hu, H.; Yang, W. *Theochem-J. Mol. Struct.* **2009**, *898*, 17–30.

(3) Sumowski, C. V.; Ochsenfeld, C. *J. Phys. Chem. A* **2009**, *113*, 11734–11741.

(4) Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 11793–11800.

(5) Liao, R.-Z.; Thiel, W. *Journal of Computational Chemistry* **2013**, *34*, 2389–2397.

(6) Hu, L.; Sderhjelm, P.; Ryde, U. *Journal of Chemical Theory and Computation* **2013**, *9*, 640–649.

(7) Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. *J. Phys. Chem. B* **2013**,

(8) Carter, E.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chemical Physics Letters* **1989**, *156*, 472 – 477.

(9) Straatsma, T.; McCammon, J. *Journal of Chemical Physics* **1991**, *95*, 1175–1188.

(10) Kelly, E.; Seth, M.; Ziegler, T. *Journal Of Physical Chemistry A* **2004**, *108*, 2167–2180.

(11) Fattebert, J.-L.; Law, R.; Bennion, B.; Lau, E.; Schwegler, E.; Lightstone, F. *J. Chem. Theory Comput.* **2009**, *5*, 2257–2264.

(12) Fattebert, J.-L.; Gygi, F. *Comput. Phys. Commun.* **2004**, *162*, 24–36.

(13) Fattebert, J.-L.; Gygi, F. *Phys. Rev. B* **2006**, *73*, 115124.

(14) Sprik, M.; Ciccotti, G. *The Journal of Chemical Physics* **1998**, *109*, 7737–7744.

(15) Dvir, H.; Silman, I.; Harel, M.; Rosenberry, T. L.; Sussman, J. L. *Chemico-Biological Interactions* **2010**, *187*, 10–22, 10th International Meeting on Cholinesterases.

(16) Wilson, I. B.; Harrison, M. A. *Journal of Biological Chemistry* **1961**, *236*, 2292–2295.

(17) Kryger, G.; Harel, M.; Giles, K.; Toker, L.; Velan, B.; Lazar, A.; Kronman, C.; Barak, D.; Ariel, N.; Shafferman, A.; Silman, I.; Sussman, J. L. *Acta Crystallogr D Biol Crystallogr* **2000**, *56 (Pt 11)*, 1385–94.

(18) Bennion, B. J.; Lau, E. Y.; Fattebert, J.-L.; Huang, P.; Schwegler, E.; Corning, W.; Lightstone, F. C. *Military Medical Science Letters* **2013**, *82*, 102–14.

(19) Bennion, B. J.; Essiz, S. G.; Lau, E. Y.; Fattebert, J.-L.; Emigh, A.; Lightstone, F. C. *PLoS One* **2015**, *10*, e0121092.

(20) Quinn, D. M. *Chemical Reviews* **1987**, *87*, 955–979.

(21) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(22) MacKerell, A. D. et al. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.

(23) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *Journal of Computational Chemistry* **1998**, *19*, 1639–1662.

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(25) Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(26) Kleinman, L.; Bylander, D. *Phys. Rev. Lett.* **1982**, *48*, 1425–1428.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(28) Froede, H. C.; Wilson, I. B. *Journal of Biological Chemistry* **1984**, *259*, 11010–11013.

(29) Silman, I.; Millard, C.; Ordentlich, A.; Greenblatt, H.; Harel, M.; Barak, D.; Shafferman, A.; Sussman, J. *Chem Biol Interact.* **1999**, *119-120*, 43–52.

(30) Ryckaert, J.; Ciccotti, G.; Berendsen, H. *Journal Of Computational Physics* **1977**, *23*, 327–341.

(31) Meijer, E. J.; Sprik, M. *Journal of the American Chemical Society* **1998**, *120*, 6345–6355.

(32) Zhang, Y.; Kua, J.; McCammon, J. A. *Journal of the American Chemical Society* **2002**, *124*, 10572–10577.

(33) Zhou, Y.; Wang, S.; Zhang, Y. *The Journal of Physical Chemistry B* **2010**, *114*, 8817–8825.

(34) Millard, C. B.; Koellner, G.; Ordentlich, A.; Shafferman, A.; Silman, I.; Sussman, J. L. *Journal of the American Chemical Society* **1999**, *121*, 9883–9884.

(35) Millard, C. B.; Kryger, G.; Ordentlich, A.; Greenblatt, H. M.; Harel, M.; Raves, M. L.; Segall, Y.; Barak, D.; Shafferman, A.; Silman, I.; Sussman, J. L. *Biochemistry* **1999**, *38*, 7032–7039.

(36) Fuxreiter, M.; Warshel, A. *Journal of the American Chemical Society* **1998**, *120*, 183–194.

(37) Zhang, Y.; Kua, J.; McCammon, J. A. *The Journal of Physical Chemistry B* **2003**, *107*, 4459–4463.

(38) Tougu, V. *Current Medicinal Chemistry -Central Nervous System Agent* **2001**, *1*, 155–170.

(39) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *The Journal of Physical Chemistry A* **2005**, *109*, 2012–2018.

(40) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. *Chemical Reviews* **2012**, *112*, 289–320.

(41) Zhao, Y.; Truhlar, D. G. *Accounts of Chemical Research* **2008**, *41*, 157–167, PMID: 18186612.

(42) Adamo, C.; Barone, V. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.

(43) Bachovchin, W. W. *Magnetic Resonance in Chemistry* **2001**, *39*, S199–S213.

(44) Rao, L.; Ke, H.; Fu, G.; Xu, X.; Yan, Y. *Journal of Chemical Theory and Computation* **2009**, *5*, 86–96.